

Chapter 416

Wilcoxon Signed-Rank Tests for Non-Inferiority

Introduction

This procedure computes power and sample size for non-inferiority tests in one-sample designs in which the one-sample t -test assumptions are violated and the *Wilcoxon Signed-Rank Test* is used.

Paired Designs

Paired data may occur because two measurements are made on the same subject or because measurements are made on two subjects that have been matched according to other variables. Hypothesis tests on paired data can be analyzed by considering the difference between the paired items as the response. The distribution of differences is usually symmetric. In fact, the distribution must be symmetric if the individual distributions of the two items are identical. Hence, the paired t -test is appropriate for paired data even when the distributions of the individual items are not normal.

In paired designs, the variable of interest is the difference between two individual measurements. Although the non-inferiority hypothesis refers to the difference between two individual means, the actual values of those means are not needed. All that is needed is their difference.

The Statistical Hypotheses

Both non-inferiority and superiority tests are examples of directional (one-sided) tests and their power and sample size could be calculated using the Wilcoxon Signed-Rank Tests procedure. However, at the urging of our users, we have developed this procedure which provides the input and output options that are convenient for non-inferiority tests. This section will review the specifics of non-inferiority testing.

Remember that in the usual t -test setting, the null (H_0) and alternative (H_1) hypotheses for one-sided upper-tail tests are defined as

$$H_0: \mu \leq \mu_0 \quad \text{versus} \quad H_1: \mu > \mu_0$$

Rejecting H_0 implies that the mean is larger than the value μ_0 . This test is called an *upper-tail test* because H_0 is rejected in samples in which the sample mean is larger than μ_0 .

The *lower-tail test* is

$$H_0: \mu \geq \mu_0 \quad \text{versus} \quad H_1: \mu < \mu_0$$

Non-inferiority tests are special cases of the above directional tests. It will be convenient to adopt the following specialized notation for the discussion of these tests.

Wilcoxon Signed-Rank Tests for Non-Inferiority

Parameter	PASS Input/Output	Interpretation
μ	μ	<i>Population mean.</i> If the data are paired differences, this is the mean of those differences. This parameter will be estimated by the study.
μ_1	μ_1	<i>Actual population mean at which power is calculated.</i> This is the assumed population mean used in all calculations.
μ_0	μ_0	<i>Non-Inferiority Mean.</i> This is the smallest (or largest) value of the mean for which the new treatment will still be considered non-inferior to the reference.
μ_R	μ_R	<i>Reference value.</i> Usually, this is the mean of a reference population. If the data are paired differences, this is the hypothesized value of the mean difference.
M_{NI}	NIM	<i>Margin of non-inferiority.</i> This is a tolerance value that defines the magnitude of difference that is not of practical importance. This may be thought of as the largest difference from the reference value that is considered to be trivial. This value is assumed to be a positive number.

Non-Inferiority Tests

A *non-inferiority test* tests that the mean is not worse than that of the baseline (reference) population by more than a small non-inferiority margin. The actual direction of the hypothesis depends on the whether higher values of the response are good or bad.

Case 1: High Values Good

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the mean of the treatment group is no less than a small amount below the reference value. The value of μ_1 at which power is calculated is often set to zero for paired tests.

Equivalent sets of the null and alternative hypotheses are

$$\begin{aligned}
 H_0: \mu \leq \mu_0 & \quad \text{versus} \quad H_1: \mu > \mu_0 \\
 H_0: \mu \leq \mu_R - M_{NI} & \quad \text{versus} \quad H_1: \mu > \mu_R - M_{NI} \\
 H_0: \mu - \mu_R \leq -M_{NI} & \quad \text{versus} \quad H_1: \mu - \mu_R > -M_{NI}
 \end{aligned}$$

Case 2: High Values Bad

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the mean of the treatment group is no more than a small amount above the reference value. The value of μ_1 at which power is calculated is often set to zero for paired tests.

Equivalent sets of the null and alternative hypotheses are

$$\begin{aligned}
 H_0: \mu \geq \mu_0 & \quad \text{versus} \quad H_1: \mu < \mu_0 \\
 H_0: \mu \geq \mu_R + M_{NI} & \quad \text{versus} \quad H_1: \mu < \mu_R + M_{NI} \\
 H_0: \mu - \mu_R \geq M_{NI} & \quad \text{versus} \quad H_1: \mu - \mu_R < M_{NI}
 \end{aligned}$$

Example

A non-inferiority test example will set the stage for the discussion of the terminology that follows. Suppose that a test is to be conducted to determine if a new cancer treatment adversely affects the mean bone density. The adjusted mean bone density (AMBD) in the population of interest is 0.002300 gm/cm with a standard deviation of 0.000300 gm/cm. Clinicians decide that if the treatment reduces AMBD by more than 5% (0.000115 gm/cm), it poses a significant health threat.

The hypothesis of interest is whether the AMBD in the treated group is greater than $0.002300 - 0.000115 = 0.002185$. The statistical test will be set up so that if the null hypothesis that the AMBD is less than or equal to 0.002185 is rejected, the conclusion will be that the new treatment is non-inferior, at least in terms of AMBD. The value 0.000115 gm/cm is called the *margin of non-inferiority*.

Wilcoxon Signed-Rank Test Statistic

The Wilcoxon signed-rank test is a popular, nonparametric substitute for the t -test. It assumes that the data follow a symmetric distribution. The test is computed using the following steps.

1. Subtract the hypothesized mean, μ_0 , from each data value. Rank the values according to their absolute values.
2. Compute the sum of the positive ranks S_p and the sum of the negative ranks S_n . The test statistic, W_R , is the minimum of S_p and S_n .
3. Compute the mean and standard deviation of W_R using the formulas

$$\mu_{W_R} = \frac{n(n+1)}{4}$$

$$\sigma_{W_R} = \sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}$$

where t represents the number of times the i^{th} value occurs.

4. Compute the z-value using

$$z_W = \frac{W_R - \mu_{W_R}}{\sigma_{W_R}}$$

The significance of the test statistic is determined by computing the p-value using the standard normal distribution. If this p-value is less than a specified level (usually 0.05), the null hypothesis is rejected in favor of the alternative hypothesis. Otherwise, no conclusion can be reached.

Power Calculation for the Wilcoxon Signed-Rank Test

The power calculation for the Wilcoxon signed-rank test is the same as that for the one-sample t -test except that an adjustment is made to the sample size based on an assumed data distribution as described in Al-Sunduqchi and Guenther (1990). The sample size n' used in power calculations is equal to

$$n' = n/W,$$

where W is the Wilcoxon adjustment factor based on the assumed data distribution.

The adjustments are as follows:

Distribution	W
Uniform	1
Double Exponential	2/3
Logistic	9/ π^2
Normal	$\pi/3$

The power is calculated as follows for a directional alternative (one-tailed test) in which $\mu_1 > \mu_0$.

1. Find t_α such that $1 - T_{df}(t_\alpha) = \alpha$, where $T_{df}(t_\alpha)$ is the area under a central- t curve to the left of x and $df = n' - 1$.
2. Calculate: $X_1 = \mu_0 + t_\alpha \frac{\sigma}{\sqrt{n'}}$.
3. Calculate the noncentrality parameter: $\lambda = \frac{\mu_1 - \mu_0}{\frac{\sigma}{\sqrt{n'}}}$.
4. Calculate: $t_1 = \frac{X_1 - \mu_1}{\frac{\sigma}{\sqrt{n'}}} + \lambda$.
5. Power = $1 - T'_{df,\lambda}(t_1)$, where $T'_{df,\lambda}(x)$ is the area to the left of x under a noncentral- t curve with degrees of freedom df and noncentrality parameter λ .

Example 1 – Power Analysis

Suppose that a test is to be conducted to determine if a new cancer treatment adversely affects the mean bone density. The adjusted mean bone density (AMBD) in the population of interest (μ_R) is 0.002300 gm/cm with a standard deviation of 0.000300 gm/cm. Clinicians decide that if the treatment reduces AMBD by more than 5% ($M_{NI} = 0.000115$ gm/cm such that $\mu_0 = \mu_R - M_{NI} = 0.002185$), it poses a significant health threat. They also want to consider what would happen if the margin of non-inferiority is set to 2.5% (0.0000575 gm/cm such that $\mu_0 = \mu_R - M_{NI} = 0.0022425$).

Following accepted procedure, the analysis will be a non-inferiority test using the Wilcoxon signed-rank test assuming a Normal data distribution at the 0.025 significance level. Power is to be calculated assuming that the new treatment has no effect on AMBD. Several sample sizes between 20 and 300 will be analyzed. The researchers want to achieve a power of at least 90%. All numbers have been multiplied by 10000 to make the reports and plots easier to read.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	Power
Higher Means Are	Better (H1: $\mu > \mu_0$)
Data Distribution	Normal
Population Size	Infinite
Alpha	0.025
N (Sample Size)	20 40 60 80 100 150 200 300
μ_0 (Non-Inferiority Mean)	21.85 22.425
μ_1 (Actual Mean)	23
σ (Standard Deviation)	3

Wilcoxon Signed-Rank Tests for Non-Inferiority

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Solve For: [Power](#)
 Higher Means Are: Better
 Hypotheses: $H_0: \mu \leq \mu_0$ vs. $H_1: \mu > \mu_0$
 Data Distribution: Normal

Power	Sample Size N	Mean		Standard Deviation σ	Alpha
		Non-Inferiority μ_0	Actual μ_1		
0.35274	20	21.9	23	3	0.025
0.63360	40	21.9	23	3	0.025
0.81170	60	21.9	23	3	0.025
0.90968	80	21.9	23	3	0.025
0.95888	100	21.9	23	3	0.025
0.99524	150	21.9	23	3	0.025
0.99951	200	21.9	23	3	0.025
1.00000	300	21.9	23	3	0.025
0.12134	20	22.4	23	3	0.025
0.20927	40	22.4	23	3	0.025
0.29540	60	22.4	23	3	0.025
0.37811	80	22.4	23	3	0.025
0.45584	100	22.4	23	3	0.025
0.62419	150	22.4	23	3	0.025
0.74810	200	22.4	23	3	0.025
0.89804	300	22.4	23	3	0.025

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
 N The sample size, the number of subjects (or pairs) in the study.
 μ_0 The non-inferiority mean. Since higher means are better, $\mu_0 = \mu_R - \text{NIM}$, where μ_R is the baseline, standard, or reference mean and NIM is the margin of non-inferiority.
 μ_1 The actual value of the population mean at which power and sample size are calculated.
 σ The standard deviation of the response (or standard deviation of differences for paired data). It measures the variability in the population.
 Alpha The probability of rejecting a true null hypothesis.

Summary Statements

A single-group design will be used to test whether the location (distribution center) is non-inferior, with a non-inferiority margin subtracted from a reference mean totaling 21.9 ($H_0: \mu \leq 21.9$ versus $H_1: \mu > 21.9$). The comparison will be made using a one-sided, one-sample Wilcoxon Signed-Rank test (the underlying data distribution is assumed to be Normal), with a Type I error rate (α) of 0.025. The underlying distribution standard deviation is assumed to be 3. To detect a mean of 23 with a sample size of 20, the power is 0.35274.

Wilcoxon Signed-Rank Tests for Non-Inferiority

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout- Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	20	25	5
20%	40	50	10
20%	60	75	15
20%	80	100	20
20%	100	125	25
20%	150	188	38
20%	200	250	50
20%	300	375	75

Dropout Rate	The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N	The evaluable sample size at which power is computed (as entered by the user). If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
N'	The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. N' is calculated by inflating N using the formula $N' = N / (1 - DR)$, with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lohknygina, Y. (2018) pages 32-33.)
D	The expected number of dropouts. $D = N' - N$.

Dropout Summary Statements

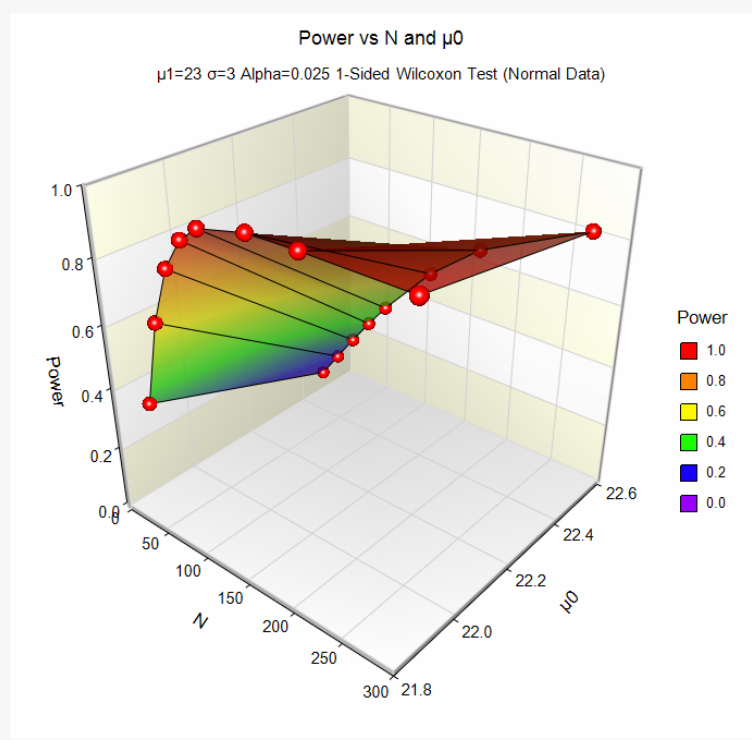
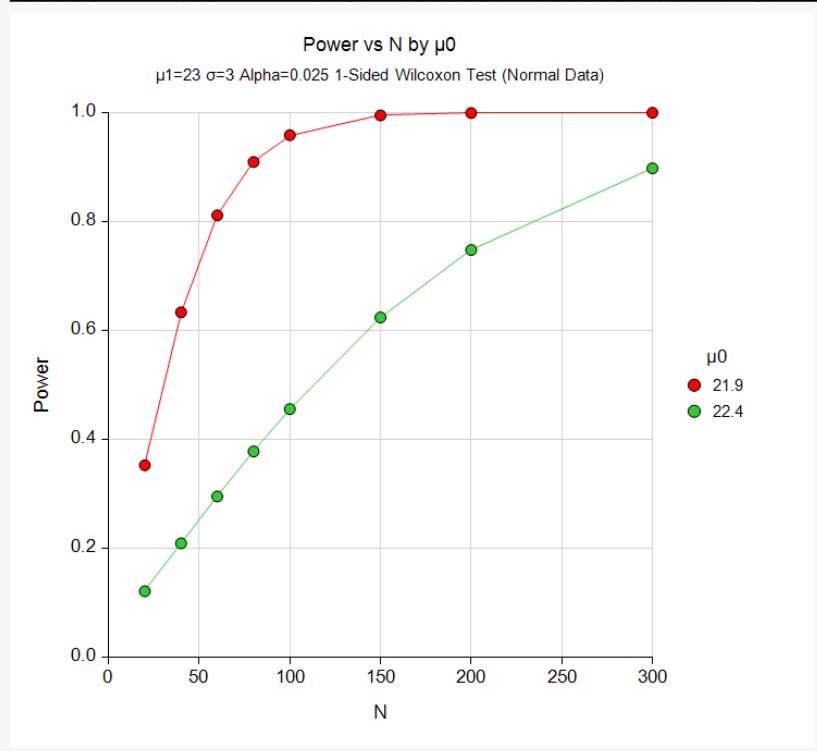
Anticipating a 20% dropout rate, 25 subjects should be enrolled to obtain a final sample size of 20 subjects.

References

- Al-Sundugchi, Mahdi S. 1990. Determining the Appropriate Sample Size for Inferences Based on the Wilcoxon Statistics. Ph.D. dissertation under the direction of William C. Guenther, Dept. of Statistics, University of Wyoming, Laramie, Wyoming.
- Chow, S.C., Shao, J., Wang, H., and Lohknygina, Y. 2018. Sample Size Calculations in Clinical Research, Third Edition. Taylor & Francis/CRC. Boca Raton, Florida.
- Julious, Steven A. 2004. 'Tutorial in Biostatistics. Sample sizes for clinical trials with Normal data.' *Statistics in Medicine*, 23:1921-1986.

Wilcoxon Signed-Rank Tests for Non-Inferiority

Plots



The above report shows that for $\mu_0 = 21.85$ (NIM = 1.15), the sample size necessary to obtain 90% power is just under 80. However, if $\mu_0 = 22.425$ (NIM = 0.575), the required sample size is about 300.

Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to know the exact sample size for each value of NIM.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab	
Solve For	Sample Size
Higher Means Are	Better (H1: $\mu > \mu_0$)
Data Distribution	Normal
Population Size	Infinite
Power.....	0.90
Alpha.....	0.025
μ_0 (Non-Inferiority Mean).....	21.85 22.425
μ_1 (Actual Mean)	23
σ (Standard Deviation).....	3

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results						
Solve For:	Sample Size					
Higher Means Are:	Better					
Hypotheses:	H0: $\mu \leq \mu_0$ vs. H1: $\mu > \mu_0$					
Data Distribution:	Normal					
Power	Sample Size N	Mean		Standard Deviation σ	Alpha	
		Non-Inferiority μ_0	Actual μ_1			
0.90215	78	21.9	23	3	0.025	
0.90005	302	22.4	23	3	0.025	

This report shows the exact sample size requirement for each value of μ_0 .

Example 3 – Validation using Chow, Shao, Wang, and Lokhnygina (2018)

Chow, Shao, Wang, and Lokhnygina (2018) page 46 has an example of a sample size calculation for a non-inferiority *t*-test where $\mu_R = 1.5$ and $M_{NI} = 0.5$. Their example obtains a sample size of 8 when $\mu_1 = 2$, $\mu_0 = 1$, $\sigma = 1$, Alpha = 0.05, and Power = 0.80.

The Wilcoxon Signed-Rank test power calculations are the same as the one-sample *t*-test except for an adjustment factor for the assumed data distribution. If we assume a uniform data distribution, we should get the same value of $N = 8$. If we assume a Normal data distribution, then the expected sample size would be $N = 8 \times \pi/3 = 9$ after rounding up.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3a** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For **Sample Size**
 Higher Means Are **Better (H1: $\mu > \mu_0$)**
 Data Distribution **Uniform**
 Population Size **Infinite**
 Power **0.80**
 Alpha **0.05**
 μ_0 (Non-Inferiority Mean) **1**
 μ_1 (Actual Mean) **2**
 σ (Standard Deviation) **1**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Solve For: **Sample Size**
 Higher Means Are: **Better**
 Hypotheses: **H0: $\mu \leq \mu_0$ vs. H1: $\mu > \mu_0$**
 Data Distribution: **Uniform**

Power	Sample Size N	Mean		Standard Deviation σ	Alpha
		Non-Inferiority μ_0	Actual μ_1		
0.81502	8	1	2	1	0.05

PASS also obtains a sample size of 8 with the uniform distribution.

Wilcoxon Signed-Rank Tests for Non-Inferiority

If we now assume a Normal data distribution and solve for sample size (**Example 3b** settings file), the results match our expected outcome.

Numeric Results

Solve For: [Sample Size](#)
 Higher Means Are: Better
 Hypotheses: $H_0: \mu \leq \mu_0$ vs. $H_1: \mu > \mu_0$
 Data Distribution: Normal

Power	Sample Size N	Mean		Standard Deviation σ	Alpha
		Non-Inferiority μ_0	Actual μ_1		
0.81502	9	1	2	1	0.05

The sample size of 9 matches the expected result.